Statistiques

I) Série statistique simple :

1) Définition :

Une série statistique est définie par la donnée d'un tableau statistique :

- D'une population déterminée
- D'un caractère (x_i)
- De l'effectif (n_i) (ou de la fréquence (F_i)) de chaque valeur (ou classe ou modalité) du caractère

Donc l'ensemble des couples (x_i , n_i) (ou (x_i , F_i)) s'appelle une série statistique à un seul caractère.

série statistique
- 1) Série statistique à caractère …………………………….. : (mesurable)
  - a) Série statistique à caractère ……….……………..… (ou regroupé par classe)
  - b) Série statistique à caractère…………………………….. (ou discontinue)
- 2) Série statistique à caractère …………………………..:(non mesurable)
- 3) Série chronologique

a) Série statistique à caractère quantitatif :

❖ **Série statistique à caractère quantitatif regroupé par classe (ou continue) :**

Exemple : Le tableau ci-dessous donne les moyennes en mathématiques de 35 élèves.

Moyenne (x_i)	[0 ,4[	[4 ,8[	[8 ,12[	[12 ,16[	[16 ,20[
Effectifs (n_i)	3	5	12	10	5

❖ **Série statistique à caractère quantitatif discret (ou discontinue)**

Exemple : Le tableau ci-dessous représente le nombre d'enfants par famille dans un quartier de 112 familles

Nombre d'enfants (x_i)	0	1	2	3	4	5	6
Effectifs (n_i)	6	20	40	24	15	5	2

b) Série statistique à caractère qualitatif :

Exemple : Le tableau suivant est une série statistique à caractère qualitatif**.** Le tableau ci-dessous donne indique la production mondiale d'huile d'olive en 2002/2003 (en millier de tonnes).

Pays de monde	Portugal	Grèce	Turque	Tunisie	Espagne	Italie	Maroc	Syrie	Reste du monde
Effectifs (n_i)	36	468	252	72	1584	720	72	252	144

c) Série chronologique :

Définition : Une série chronologique est une série de valeurs provenant d'une même variable observée à des instants régulièrement espacés dans le temps.

Exemple : Le tableau ci-dessous indique les résultats du baccalauréat de l'année2000 Jusqu'à l'année 2005 dans un établissement A.

Année (x_i)	2000	2001	2002	2003	2004	2005
Nombre de réussites (n_i)	350	767	840	590	880	1000

1) *Le coefficient multiplicateur C qui permet de passer de l'année b à l'année n est égal à .*

$$C = \frac{valeur\ de\ l'année\ \mathrm{n}}{valeur\ de\ l'année\ \mathrm{b}}$$

2) *L 'indice de l'année n , base 100 en l'année b est égal à :* $I = C \times 100 = \frac{valeur\ de\ l'année\ \mathrm{n}}{valeur\ de\ l'année\ \mathrm{b}} \times 100$

Retenons :

- **Population :** ……………………………………………………………..
- **Individu :** ………………………………………………………………….
- **Caractère(ou variable) en statistique** (x_i)

(x_i) Caractère

- Caractère quantitatif : (mesurable)
  - Caractère quantitatif discret (ou discontinue) (x_i est une valeur)
  - ou
  - Un caractère quantitatif continue (regroupé par classe) ; (x_i est une classe)
- ou
- Caractère qualitatif : (non mesurable) ; (x_i est une modalité)

- **Effectif** :(noté (n_i))

L'effectif est le nombre d'individus statistique correspondant a une valeur (ou une classe ou une modalité)

- **Fréquence** : (noté (F_i))

$F_i = \frac{l'effectif}{l'effectif \cdot totall} = \frac{n_i}{N}$ avec N =la somme des effectifs (effectif total) et n_i l'effectif de x_i

- **Pourcentages** : (noté (p_i))

$$p_i = F_i \times 100 = \frac{n_i}{N} \times 100$$

- **Effectifs cumulés :** Effectifs cumulés $\begin{cases} \text{Effectifs cumulés croissantes noté } n_i \nearrow \\ \text{Effectifs cumulés décroissantes noté } n_i \searrow \end{cases}$

* On appelle effectif cumulé croissant de la valeur (x_i) le nombre d'individus ayant des valeurs du caractère Inférieur ou égale à (x_i)

* On appelle effectif cumulé décroissant de la valeur (x_i) le nombre d'individus ayant des valeurs du caractère supérieur ou égale à (x_i)

- **Fréquences cumulés :** Fréquences cumulées $\begin{cases} \text{Fréquences cumulés croissantes noté } F_i \nearrow \\ \text{Fréquences cumulés décroissantes noté } F_i \searrow \end{cases}$

Pour calculer les fréquences cumulés (croissantes ou décroissantes) il faut calculer la fréquence de chaque variable (x_i) $F_i = \frac{n_i}{N}$ **avec N =** la somme des effectifs (l'effectif total)

* On appelle fréquences cumulés croissant de la valeur (x_i) la somme des fréquences ayant des valeurs du caractère Inférieur ou égale à (x_i)
* On appelle fréquences cumulés décroissant de la valeur (x_i) la somme des fréquences ayant des valeurs du caractère supérieur ou égale à (x_i)

a) Effectif, fréquence, pourcentage :

Activité N°1 : *Le tableau ci-dessous donne la répartition des élèves d'une classe selon leurs âges.*

âges (x_i)	**14**	**15**	**16**	**17**	**18**
Effectifs (n_i)	**4**	**15**	**6**	**10**	**5**

1) *Quelle est le type de cette série ?*..

2) ***Quelle est la population étudiée ?***

……………………………………………………………………………….

……………………………………………………………………………….

Calculer les fréquences et les fréq.uences en pourcentage ?

Compléter le tableau suivant :

Moyenne (x_i)	***14***	***15***	***16***	***17***	***18***	***Total***
Effectifs (n_i)	***4***	***15***	***6***	***10***	***5***	
Fréquence (F_i)	$F_1=\frac{n_1}{N}=\frac{4}{40}=0,1$					
Pourcentages (p_i)	$0,1\times100=$ ***10***					

Remarques :

* *La somme des fréquences est égale à1* $F=F_1+F_2+.....+F_p=1$
* *La somme des Pourcentages est égale à100* $p_1+p_2+.....+p_i=100$

b) Effectifs cumulés –Fréquence cumulés :

Lorsque on étudie un caractère quantitatif discret. Pour calculer l'effectif cumulés ou fréquence cumulés il faut classer les valeurs de la variable (x_i) *a l'ordre croissantes lorsque on étudie un caractère quantitatif discret.*

i) Effectifs cumulés :

Exemple :

Le tableau ci-dessous représente le nombre d'enfants par famille dans un quartier de 110 familles

Nombre d'enfants (x_i)	***0***	***1***	***2***	***3***	***4***	***5***	***6***
Effectifs (n_i)	***n_1=6***	***n_2=20***	***n_3=40***	***24***	***15***	***3***	***2***

Donner les effectifs cumulés croissants et les effectifs cumulés décroissants de cette série ?

Compléter ce tableau :

Nombre d'enfants (x_i)	***0***	***1***	***2***	***3***	***4***	***5***	***6***
Effectifs (n_i)	***n_1=6***	***n_2=20***	***n_3=40***	***24***	***15***	***3***	***2***
Effectifs cumulés croissants	$n_1 \square =6$						
Effectifs cumulés décroissants	***110***						***2***

ii) Fréquences cumulés :

Exemple : *Soit **le** tableau ci-dessous donne la moyenne annuelle de 64 élèves de 2^{éme} année d'un établissement scolaire.*

Moyenne (x_i)	***[0 ,4[***	***[4 ,8[***	***[8 ,12[***	***[12 ,16[***	***[16 ,20[***
Effectifs (n_i)	***6***	***12***	***24***	***14***	***8***

1) *Quelle est le type de cette série ?*..

2) ***Quelle est la population étudiée ?***

………………………………………………………………………………………………………

……………………………………………………………………………………………………....

……………………………………………………………………………………………………...

3) ***Donner les fréquences cumulées croissantes et les fréquences cumulées décroissants ?***

Compléter ce tableau

Moyenne (x_i)	***[0 ,4[***	***[4 ,8[***	***[8 ,12[***	***[12 ,16[***	***[16 ,20[***	***Total***
Effectifs (n_i)	***n_1= 6***	***n_2=12***	***24***	***14***	***8***	
Fréquences F_i	$F_1 = \frac{6}{64} = 0,09375$					
fréquences cumulées croissantes	$F_1\nearrow = F_1 =$ $0,09375$					
fréquences cumulées décroissantes						

2) Paramètres d'un caractère statistique :

On distingue deux types de paramètres sur un caractère statistique. *Les* parmaètres d'un caractère statistique $\begin{cases} \text{Les paramétres de position} \\ \text{les paramétres de dispersion} \end{cases}$

i) Paramètres de position :

Les parmaètres de position $\begin{cases} \text{..........................} \\ \text{..........................} \\ \text{..........................} \\ \text{..........................} \end{cases}$

► Mode d'une série statistique :

On note (M_O)

Le mode est une valeur du caractère x_i (ou modalité) qui correspond à…………………………………………

Pour un caractère quantitatif

❖ **Caractère discret**

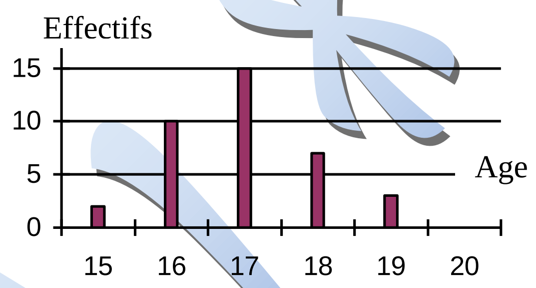
Exemple 1: Le tableau ci-dessous représente la taille en m des joueurs d'une équipe de basket-ball.

Taille (x_i)	**1 ,75**	**1,85**	**1,90**	**1,95**	**2,00**
Effectifs (n_i)	**2**	**3**	**4**	**3**	**1**

Quelle est le mode de cette série ?

………

Exemple 2 : Ce diagramme en bâtons suivant donne la repartions des âges des élevés d'une classe de 3^{éme} économie.



∗ Déterminer le mode de cette série ?

……………………………………………………
……………………………………………………
…………………………………

❖ **Caractère continue**

La classe modale est la classe dont l'effectif est relativement le plus élevé et on attribut au mode la valeur centrale de cette classe

Exemple3 :

Le tableau ci-dessous donne les moyennes en mathématiques de 35 élèves.

Moyenne (x_i)	**[0 ,4[**	**[4 ,8[**	**[8 ,12[**	**[12 ,16[**	**[16 ,20[**
Effectifs (n_i)	**3**	**5**	**12**	**10**	**5**

Déterminer la classe modale et le mode de cette série ?

………

Pour un caractère qualitatif

Exemple1 : On demande a 40 personnes leur couleur préférée parmi les suivantes : rouge, bleu, vert, jaune, noir, blanc, on a les résultat suivants :

Couleur	**rouge**	**bleu**	**vert**	**jaune**	**noir**	**blanc**
Effectifs (n_i)	**5**	**4**	**10**	**8**	**6**	**7**

∗ déterminer le mode de cette série ?

………

Remarques :

- Une série statistique peut présenter plusieurs modes. la série est dite multi mode
- Une série statistique qui n'a qu un seul mode est dite uni modale
- Une série statistique qui n'a que deux modes est dite bimodale

▶ **Médiane d'une série statistique :** On note (M_e)

La médiane est le nombre qui sépare la série en deux groupes de même effectif.
On considère une série statistique à caractère quantitatif

a) Série statistique à caractère discret

Pour déterminer la médiane d'une série statistique à caractère discret, on classe les individus suivant les valeurs du caractère dans l'ordre croissant (ou décroissant)

- *Si l'effectif total N est impair la médiane est la valeur du caractère de l'individu de rang* $\frac{N+1}{2}$
- *Si l'effectif total N est pair la médiane est la valeur moyenne des deux valeurs du caractère des deux individus de rang* $\frac{N}{2}$ *et* $\frac{N}{2}+1$

Exemple 1:

Le tableau ci-dessous représente le poids en kg des joueurs d'une équipe de hand-ball .

Poids (x_i)	**75**	**80**	**85**	**90**	**95**
Effectifs (n_i)	**1**	**4**	**3**	**2**	**1**

75 ,80,80,80,80,85,85,85,90,90,95

N =…….. est ………………..alors la valeur du caractère de l'individu de rang
…………………… est…………donc la médiane M_e = ...

Exemple 2 :

Le tableau ci-dessous représente le nombre d'enfants par famille dans un quartier de 14 familles

Nombre d'enfants (x_i)	**0**	**1**	**2**	**3**	**4**
Effectifs (n_i)	**2**	**5**	**4**	**2**	**1**

0 ,0,1,1,1,1,1,2,2,2,2 , 3,3,4

N =……..est …………….. alors la valeur moyenne des deux valeurs du caractère des deux individus de rang …………… et ………………. est ……………….. .Donc la médiane M_e = ...

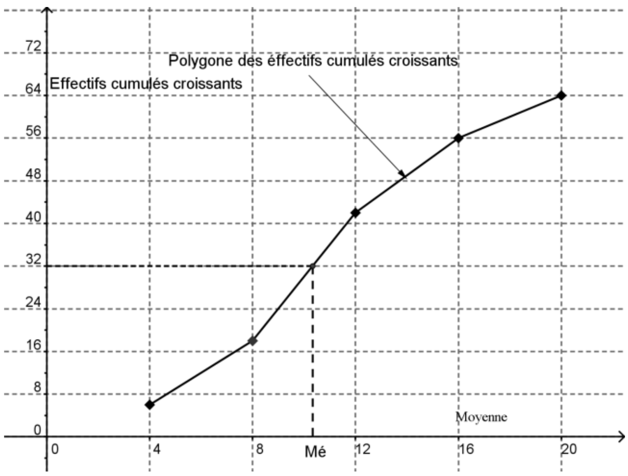
b) Série statistique à caractère continu

Méthode graphique pour déterminer la médiane :

- *Le calcul de la médiane nécessite la détermination des effectifs cumulés ou des fréquences cumulés (croissantes ou décroissantes)*
- *Dans le cas ou le caractère est connu par son polygone des effectifs cumulés la médiane est l'abscisse du point de la courbe dont l'ordonnée est égale à* $\frac{N}{2}$ *(ou N est l'effectif total)*

Exemple *:Soit le tableau ci-dessous donne la moyenne annuelle de 64 élèves de* $3^{iéme}$ *année d'un établissement scolaire.*

Moyenne (x_i)	**[0 ,4[**	**[4 ,8[**	**[8 ,12[**	**[12 ,16[**	**[16 ,20[**
Effectifs (n_i)	**6**	**12**	**24**	**14**	**8**
Effectifs cumulés croissants	**6**	**18**	**42**	**56**	**N=64**



Le polygone des effectifs cumulés croissants correspondant

La moitié de l'effectif total est $\frac{N}{2}=\frac{64}{2}=32$ d'après le graphique, la médiane est l'abscisse du point de la courbe dont l'ordonnée est égale à 32 donc le médiane $M_e=10$

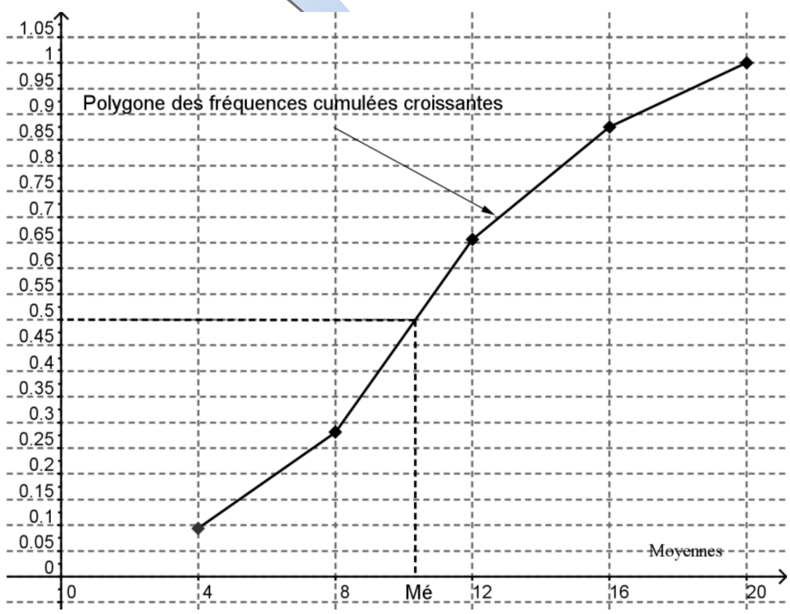
- *Dans le cas ou le caractère est connu par son polygone des fréquences cumulés (croissantes ou décroissantes) la médiane est l'abscisse du point de la courbe dont l'ordonnée est égale à* ***0,5***

<u>Exemple</u> *:*

Soit ***le*** *tableau ci-dessous donne la moyenne annuelle de 64 élèves de 3[iéme] année d'un établissement scolaire.*

Moyenne (x_i)	***[0 ,4[***	***[4 ,8[***	***[8 ,12[***	***[12 ,16[***	***[16 ,20[***	***Total***
Effectifs (n_i)	***6***	***12***	***24***	***14***	***8***	***64***
Fréquences F_i	0,09375	0,1875	0,375	0,21875	***0,125***	***F=1***
Fréquences cumulées croissantes	0,09375	***0,28125***	***0,65625***	***0,875***	***F=1***	

Le polygone des fréquences cumulées croissantes correspondant



la médiane est l'abscisse du point de la courbe dont l'ordonnée est égale à **0,5** donc $M_e \approx 10$

▶ *Moyenne d'une série statistique :*

On considère un caractère quantitatif

a) *Cas d'une série à caractère discret*

Quand la série statistique est discrète, on a un tableau du type :

Valeur du caractère (x_i)	x_1	x_2	x_3	$....x_p$	***Total***
Effectifs (n_i)	n_1	n_2	n_3	$....n_p$	***N***

On appelle moyenne de cette série le nombre : $\bar{x} = \dfrac{n_1x_1+n_2x_2+.....+n_px_p}{N} = \sum_{i=1}^{n}\dfrac{n_ix_i}{N}$

ou N : *l'effectif total* $N = n_1 + n_2 + n_3 + ... + n_p$

b) *Cas d'une série à caractère continue*

Quand la série statistique est ***continue****, on a un tableau du type :*

Classe du caractère (x_i)	$[a_1\ ,\ a_2[$	$[a_2\ ,\ a_3[$	$[a_3\ ,\ a_4[$	$....[a_p\ ,\ a_{p+1}[$
Centre de l'intervalle	$c_1 = \dfrac{a_1+a_2}{2}$	$c_2 = \dfrac{a_2+a_3}{2}$	$c_3 = \dfrac{a_3+a_4}{2}$	$...c_p = \dfrac{a_p+a_{p+1}}{2}$
Effectifs (n_i)	n_1	n_2	n_3	$....n_p$

Pour calculer la moyenne d'une telle série, on utilise la formule précédente en remplaçant x_i *par le centre* c_i *de l'intervalle* $[a_i\ ,\ a_{i+1}[$ *. La moyenne de cette série est alors:* $\bar{x} = \dfrac{n_1c_1+n_2c_2+.....+n_pc_p}{N} = \sum_{i=1}^{n}\dfrac{n_ic_i}{N}$ **ou** $c_i = \dfrac{a_i+a_{i+1}}{2}$

▶ *Quartiles :*

La médiane partage la série en deux groupes de même effectif. Les quartiles partagent la série en quarte groupes de même effectif.

Ils sont donc au nombre de trois Q_1 , Q_2 *et* Q_3 *ou* Q_2 *est la médiane, les intervalles qu ils définissent contiennent chacun 25% des observations, soit un quart de l'effectif,* $\dfrac{N}{4}$ *comme le montre le schéma :*

	50%		50%	
Population	*25%*	*25%*	*25%*	*25%*
Caractère	x_{min}	Q_1	$M_é$	Q_3	x_{max}

a) *Cas d'une série à caractère discret*

- ❖ Q_1 **:** *La valeur du caractère de rang* $\dfrac{N}{4}$
- ❖ Q_2 **:** $M_é$
- ❖ Q_3 **:** *La valeur du caractère de rang* $\dfrac{3N}{4}$
- ❖ *L'intervalle* $[Q_1\ \ Q_3]$ *s'appelle intervalle interquartile*
- ❖ *Le réel* $Q_3 - Q_1$ *s'appelle l'écart interquartile*

Remarque :

Si N n'est pas multiple de 4 alors Q_1 *(respectivement* Q_3 *) est la valeur du caractère dont le rang est immédiatement supérieur à* $\dfrac{N}{4}$ *(respectivement* $\dfrac{3N}{4}$ *)*

b) *Cas d'une série à caractère continue :*

Méthode graphique : Le procédé de détermination des quartiles est identique à celui de détermination des médianes

- Q_1 *est l'abscisse du point de la courbe dont l'ordonnée est égale à* $\frac{N}{4}$ *sur le polygone des effectifs cumulés et* $\frac{1}{4}$ *sur le polygone des fréquences cumulées*
- $Q_2 = M_é$
- Q_3 *est l'abscisse du point de la courbe dont l'ordonnée est égale à* $\frac{3N}{4}$ *sur le polygone des effectifs cumulés et* $\frac{3}{4}$ *sur le polygone des fréquences cumulées.*

2[iéme] méthode: Au moyen d'une interpolation linéaire en utilisant le tableau des effectifs cumulés.

i) Paramètres de dispersion :

Les parmaètres de dispersion $\begin{cases}\text{L'étendue} \\ \text{L'écart type} \\ \text{Variance}\end{cases}$

- **Etendue :** *L'étendue d'une série statistique est la différence entre ses deux valeurs extrêmes (la plus grande et la plus petite valeur du caractère)*

Exemple 1 : *Le tableau ci-dessous donne la répartition des élèves d'une classe selon leurs âges.*

âges (x_i)	**14**	**15**	**16**	**17**	**18**
Effectifs (n_i)	**4**	**15**	**6**	**10**	**5**

L'étendue est : ………. - ….. = ………

Exemple2 : *Dans une entre prise A les salaires horaires sont classés de la façon suivantes*

Salaires en dinars (x_i)	**[1, 3[**	**[3 ,4[**	**[4, 5[**	**[5 ,7[**	**[7, 9[**
Effectifs (n_i)	**50**	**65**	**17**	**10**	**8**

L'étendue est : ………………

- **Variance :**

Cas d'une série à caractère discret

Quand la série statistique est discrète, on a un tableau du type :

Valeur du caractère (x_i)	x_1	x_2	x_3	**….** x_p	***Total***
Effectifs (n_i)	n_1	n_2	n_3	**….** n_p	***N***

La variance de la série est le nombre noté V et défini par

$$V = \frac{n_1(x_1)^2 + n_2(x_2)^2 + \dots + n_p(x_p)^2}{N} - (\overline{x})^2$$

ou N : *l'effectif total* **N** $= n_1 + n_2 + n_3 + \dots + n_p$

Cas d'une série à caractère continue

Quand la série statistique est ***continue****, on a un tableau du type :*

Classe du caractère (x_i)	$[a_1, a_2[$	$[a_2, a_3[$	$[a_3, a_4[$	$\dots [a_p, a_{p+1}[$
centre de l'intervalle	$c_1 = \frac{a_1 + a_2}{2}$	$c_2 = \frac{a_2 + a_3}{2}$	$c_3 = \frac{a_3 + a_4}{2}$	$\dots c_p = \frac{a_p + a_{p+1}}{2}$
Effectifs (n_i)	n_1	n_2	n_3	**….** n_p

La variance de la série est le nombre noté V et défini par

$$V = \frac{n_1(c_1)^2 + n_2(c_2)^2 + \dots + n_p(c_p)^2}{N} - (\overline{x})^2$$

ou N : *l'effectif total* **N** $= n_1 + n_2 + n_3 + \dots + n_p$

- ***L'écart type* :**

L'écart type de la série est le nombre noté σ et défini par $\sigma = \sqrt{V}$

- **Remarque :*On peut obtenir N, $\overline{x}$, V et σ directement à partir de la calculatrice***

Utilisation des calculatrices

Série statistique à une variable:

A titre d'exemple, on donne le mode de fonctionnement d'une calculatrice.

Le tableau ci-dessous donne la répartition des élèves d'une classe selon leurs âges.

âges (x_i)	**14**	**15**	**16**	**17**	**18**
Effectifs (n_i)	**4**	**15**	**6**	**10**	**5**

Pour choisir le mode de fonctionnement en statistique appuyer sur 2ndF MODE 1

Pour entrer les données taper

x_i	STO	n_i	M^+
14	STO	**4**	M^+
15	STO	**15**	M^+
16	STO	**6**	M^+
17	STO	**10**	M^+
18	STO	**5**	M^+

La calculatrice affiche l'effectif total N = 40

Pour afficher la moyenne $\overline{X}$ appuyer sur :	RCL	$\overline{X}$	*On obtient :*	**$\overline{X}$ =15 ,925**
Pour afficher L'écart type σ appuyer sur :	RCL	σ_x	*On obtient :*	**σ_x =1 , 232629304**
Pour afficher la variance V appuyer sur :	RCL	σ_x x^2	*On obtient :*	**V= 1,519375**
Pour afficher l'effectif total N appuyer sur :	RCL	n	*On obtient :*	**N = 40**

4) *Représentation graphique d'une série statistique :*

Il y a quatre types de représentation graphique d'une série statistique

Représentation graphique d'une série statistique →

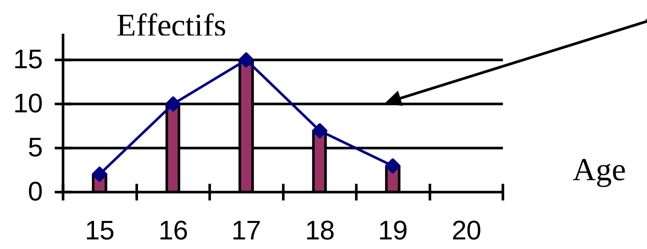
- *1) Diagrammes en bâtons*
- *2) Histogrammes*
- *3) Diagrammes à secteurs circulaires*
- *4) Diagrammes en boite ou boite de dispersion*

i) *Diagrammes en bâtons* : *Ce type de présentation graphique est en général, utilisé lorsque on considère des séries statistiques à caractère quantitatif discret (ou discontinue)*

Exemple :

Ce diagramme en bâtons suivant donne la repartions des âges des élevés d'une classe de 3[éme] économie.

Polygone des effectifs



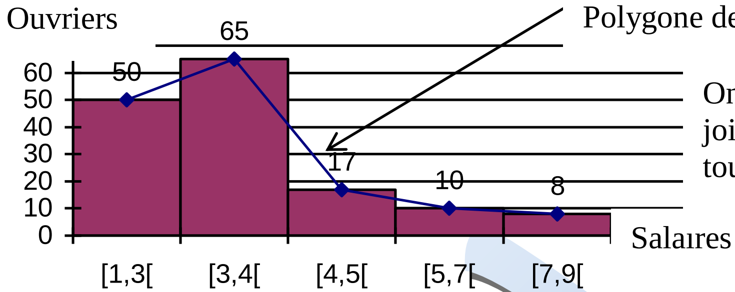
Remarque : chaque segment du diagramme à une extrémité non située sur l'axe des abscisses. En joignant ces extrémités on forme le polygone des effectifs

ii) <u>Histogrammes</u> : *Ce type de présentation graphique est utilisé lorsque on considère des séries statistiques à caractère quantitatif continu (ou regroupé par classe).Il s'obtient en formant des rectangles dont les bases sont les amplitudes des différents intervalles.*

<u>Exemple</u> : *Dans une entre prise A les salaires horaires sont classés de la façon suivantes*

Salaires en dinars (x_i)	***[1, 3[***	***[3 ,4[***	***[4, 5[***	***[5 ,7[***	***[7, 9[***
Effectifs (n_i)	***50***	***65***	***17***	***10***	***8***

Cet histogramme donne la repartions des ouvriers suivant leurs salaires horaires.



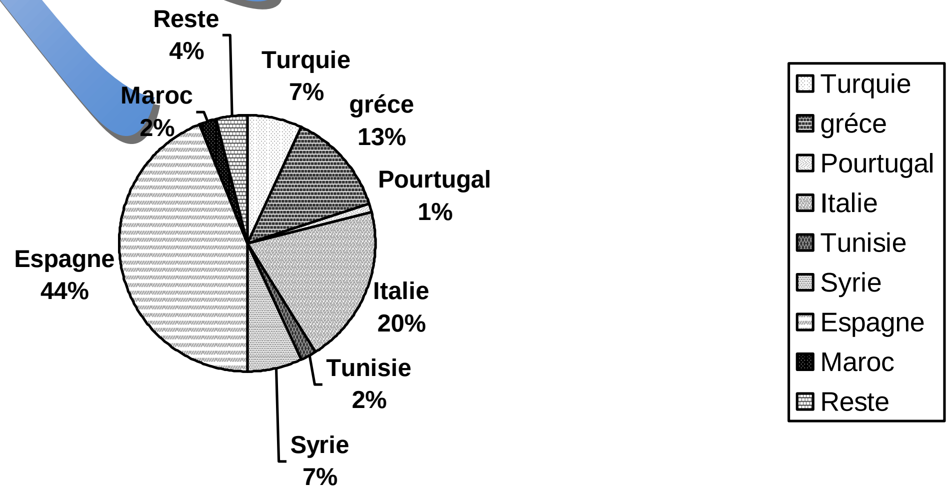
On obtient le polygone des effectifs en joignant les milieux des cotés supérieurs de tous les rectangles

iii) <u>Diagrammes à secteurs circulaires</u> :

*Ce type de présentation graphique est en général utilisé dans le cas ou le caractère étudier est **<u>qualitatif</u>**. Il consiste à déterminer sur un disque circulaire des secteurs dont les aires sont proportionnelles aux effectifs (ou aux fréquences).*

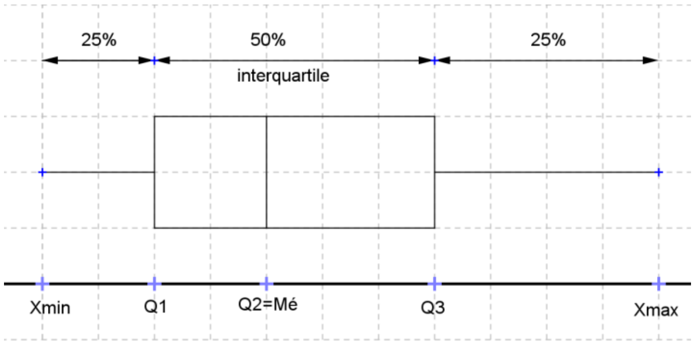
<u>Exemple</u> : *ce diagramme à secteur circulaire indique la production mondiale d'huile d'olive en 2002/2003 des principaux pays producteurs.*

Pays de monde	***Portugal***	***Grèce***	***Turque***	***Tunisie***	***Espagne***	***Italie***	***Maroc***	***Syrie***	***Reste du monde***
Effectifs (n_i)	***36***	***468***	***252***	***72***	***1584***	***720***	***72***	***252***	***144***
Fréquences F_i	0,01	0,13	***0,07***	***0 ,02***	***0,44***	***0,2***	***0 ,02***	***0,07***	***0 ,04***
Angle en degré $\alpha_i = 360 \times F_i$	***3 ,6***								



iv) Diagrammes en boite ou boite de dispersion :

Un diagramme en boite est un rectangle délimité par le premier quartile et le troisième quartile.
Pour l'obtenir, on trace un axe horizontal sur lequel on place les valeurs de Q_1, Q_3 et M_e
L'un des côtés du rectangle a pour longueur l'écart interquartile $Q_3 - Q_1$, l'autre est quelconque.
On complète ce diagramme en traçant deux traits horizontaux : l'un joignant Q_1 au minimum de la série et l'autre joignant Q_3 au maximum de la série.



Exemple :

Le tableau ci-dessous donne la répartition de 300 membres d'un club sportif, selon leur taille en cm .

Taille (en cm) (x_i)	**160**	**165**	**170**	**175**	**180**	**185**	**190**	**Total**
Effectifs (n_i)	**9**	**33**	**74**	**93**	**64**	**21**	**6**	**300**
Effectifs cumulés croissants	**9**	**42**	**116**	**209**	**273**	**294**	**N=300**	

❖ *Les quartiles de cette série :*

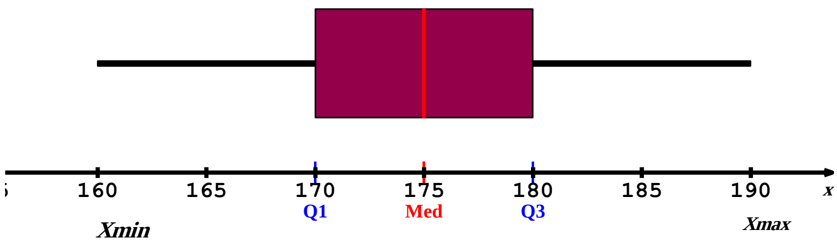
***** Q_1 **:** *La valeur du caractère de l'individu de rang* $\frac{N}{4} = \frac{300}{4} = 75$ $(n_i \,\square\,)$ *donc* $Q_1 = 170$

***** *N= 300 est pair alors la valeur moyenne des deux valeurs du caractère des deux individus de rang* $\frac{N}{2} = 150$ *et* $\frac{N}{2} + 1 = 151$ *est* $\frac{175+175}{2} = 175$ *.Donc la médiane* $Q_2 = M_é = 175$

*****Q_3 **=***La valeur du caractère de l'individu de rang* $\frac{3N}{4} = 225$ $(n_i \,\square\,)$ *donc* $Q_3 = 180$

❖ *Le diagrammes en boite ou la boite de dispersion de cette série :*

Effectif total : 300
Minimum : 160 ; Maximum : 190
Premier quartile : $Q_1 = 170$ *; Médiane :* $Q_2 = 175$
Troisième quartile : $Q_3 = 18$



Activité N°1 :

Compléter les phrases ci –dessous en utilisant l'un de ces mots .
un histogramme , les individus, Le mode, L'étendue, une population , qualitatif, quantitatif , continue , discret , un diagramme en bâton ,

1. *L'ensemble sur le quel porte une étude statistique s'appelle …………………… . Les éléments de cet ensemble sont ………………… de la population*
2. *Un caractères est dit……………… s'il n'est pas mesurable (exemples : couleur,moyen de transport….).*
3. *Un caractère est dit…………………… s'il est mesurable(exemples :Nombre d'enfant, poids, taille,…)*
4. *Graphiquement, on représente une série statistique à caractère quantitatif ………………..., par …………………… ..et on représente une série statistique à caractère quantitatif ………………..., par …………………… .*
5. *……………….est une valeur du caractère x_i (ou modalité) qui correspond à l'effectif le plus grand.*
6. *…………………d'une série statistique est la différence entre ses deux valeurs extrêmes (la plus grande et la plus petite valeur du caractère)*

Activité N°2 : ***(Cas d'une série à caractère discret)***

Le tableau suivant indique la répartition de 128 appartements d'une cité suivant le nombre de leur pièces

Nombre des pièces (x_i)	**1**	**2**	**3**	**4**	**5**	**6**
Nombre d'appartement (n_i)	**10**	**34**	**50**	**24**	**8**	**2**

1) *Trouver le mode, l'étendue et la médiane de cette série*
2) *Représenter le diagramme en bâton à effectifs de cette série*
3) *Donner les effectifs cumulés croissants de cette série*
4) *a) Calculer les fréquences cumulées croissantes de cette série*
b) représenter le polygone des fréquences cumulées croissantes
5) *Calculer la moyenne $\overline{x}$ de cette série*
6) *Calculer l'écart type σ et la variance V de cette série statistique*
7) *Déterminer les quartiles de cette série*
8) *Représenter la boite de dispersion de cette série*

Activité N°3 : ***(Cas d'une série à caractère continue)***

Un grossiste a relevé le montant des factures de livraison reçue au cours d'un mois.

Montant des factures (En dinars)	***[0,100[***	***[100,200[***	***[200,300[***	***[300,400[***	***[400,500[***	***[500,600[***
Nombre des factures	***10***	***20***	***18***	***14***	***12***	***6***

1) *Tracer le histogramme de cette série*
2) *a) Donner sa classe modale*
b) Déterminer pour cette série le centre de chaque classe.
c) Calculer la moyenne $\overline{x}$ de cette série
3) *Calculer l'écart type σ et la variance V de cette série statistique*
4) *Calculer les fréquences cumulées croissantes de cette série*
5) *a) Donner les effectifs cumulés croissants de cette série*
b) *représenter le polygone des effectifs cumulés croissants de cette série*
c) *déduire graphiquement la médiane de cette série*
6) *Calculer par la méthode d'interpolation linéaire les quartiles Q_1 , Q_2 et Q_3 de cette série*
7) *Représenter la boite de dispersion de cette série*

Activité N°4: *Le tableau suivant représente les distances des domiciles des élèves au lycée.*

Distance (km)	[0,2 [	[2,4[	[4,6[	[6,8[	[8,10[
Nbr des élèves	240	300	260	150	50

On suppose que la répartition est uniforme à l'intérieur de chaque intervalle-

a) Rappel

1) Déterminer la population et le caractère étudier dans cette série.
2) Tracer l'histogramme des effectifs de cette série.
3) Compléter le tableau suivant :

Classe					
Effectif					
Eff.cumulées ↗					
Fréquence					
Fré. cumulées ↗					
Pourcentages					
Centres des Classes					

b) Paramètres de position - - Mode - Moyenne - Médiane - Quartiles

1) Calculer l'étendue de cette série.
2) Qu'elle est la classe modale de cette série.
3) Calculer la moyenne des distances (domiciles-lycée).
4) Tracer le polygone des effectifs cumulées croissantes.
5) Tracer le polygone des fréquences cumulées croissantes.
6) a-Déterminer graphiquement puis par le calcul les trois quartiles.
   b- Déterminer l'intervalle et l'écart interquartiles.
   c-Représenter le diagramme en boite de cette série.

C/ Paramètres de dispersion-variance-Ecart type.-Etendu

1) Calculer la variance et l'écart-type de cette série.
2) Combien y-a-t-il d'élèves dans l'intervalle $[\bar{X}-\sigma, \bar{X}+\sigma]$
3) Combien y-a-t-il d'élèves dans l'intervalle $[\bar{X}-2\sigma, \bar{X}+2\sigma]$
4) On donne le tableau suivant :

Distance (km)	[0,2 [	[2,4[	[4,6[	[6,8[	[8,10[
Nbr des élèves	340	400	360	250	150

a- Que remarquer vous ?
b-Déduire la moyenne ; la variance et l'écart-type de la nouvelle série.
5) même question (N°C/4) avec le tableau suivant :

Distance (km)	[0,2[	[2,4[	[4,6[	[6,8[	[8,10[
Nbr des élèves	480	600	520	300	100

Activité N°5:

Afin de centrer les lunettes en face des pupilles, les opticiens s'intéressent à « l'écartement inter pupillaire ». Il est ainsi mesuré en millimètres.
On a mesuré cet écartement, désigné par e, pour 50 femmes et les résultats statistiques sont données ci-dessous avec une répartition en classes :

e (mm)	[55 ; 56,5[	[56,5 ; 58[	[58 ; 59,5[	[59,5 ; 61[	[61 ; 62,5[	[62,5 ; 64[	[64 ; 65,5[	[65,5 ; 67[	[67 ; 68,5[	[68 ; 70[
Nombre	2	3	4	7	9	8	7	5	3	2

a) Donnez les différents indices de position et de dispersion de cette série.
b) Dessinez la boîte à moustache de cette série. Vous donnerez évidemment le détail des calculs
c) Déterminez le pourcentage des valeurs de la série comprises entre -2σ et $+ 2\sigma$.

Activité N°6 :

La série suivante donne la taille en cm des 550 nourrissons nés dans une maternité dans l'année.

Taille x_i	45	46	47	48	49	50	51	52	53	54	55	56
Nombre	20	31	52	56	65	90	80	82	38	18	9	9

a) Donnez les différents indices de position et de dispersion de cette série.
b) Dessinez la boîte à moustache de cette série. Vous donnerez évidemment le détail des calculs

Activité N°1 :

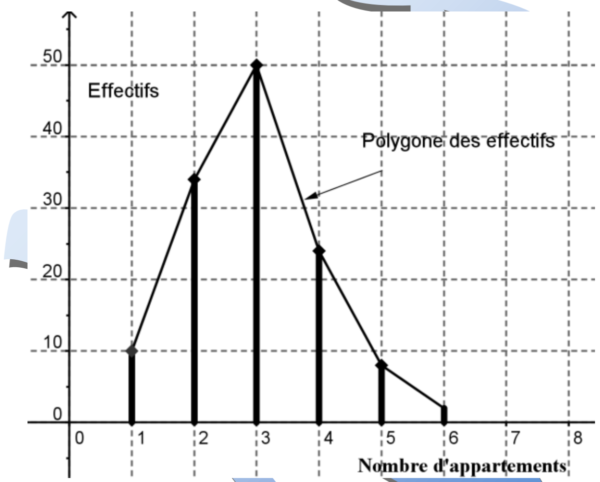
1. L'ensemble sur le quel porte une étude statistique s'appelle une population. Les éléments de cet ensemble sont les individus de la population
2. Un caractères est dit qualitatif, s'il n'est pas mesurable (exemples : couleur,moyen de transport….).
3. Un caractère est dit quantitatif s'il est mesurable (exemples : Nombre d'enfant, poids, taille,…)
4. Graphiquement, on représente une série statistique à caractère quantitatif continue, par un histogramme et on représente une série statistique à caractère quantitatif discret , par un diagramme en bâton .
5. Le mode est une valeur du caractère x_i (ou modalité) qui correspond à l'effectif le plus grand .
6. L'étendue d'une série statistique est la différence entre ses deux valeurs extrêmes (la plus grande et la plus petite valeur du caractère)

Activité N°2 : (Cas d'une série à caractère discret)

1) a) Le mode de cette série est : $M_O = 3$

- L'étendue de cette série est :6-1=5
- N =128 est pair alors la valeur moyenne des deux valeurs du caractère des deux individus de rang $\frac{N}{2} = \frac{128}{2} = 64$ et $\frac{N}{2} + 1 = 65$ Donc la médiane $M_e = \frac{3+3}{2} = 3$

2) Le diagramme en bâton à effectifs de cette série



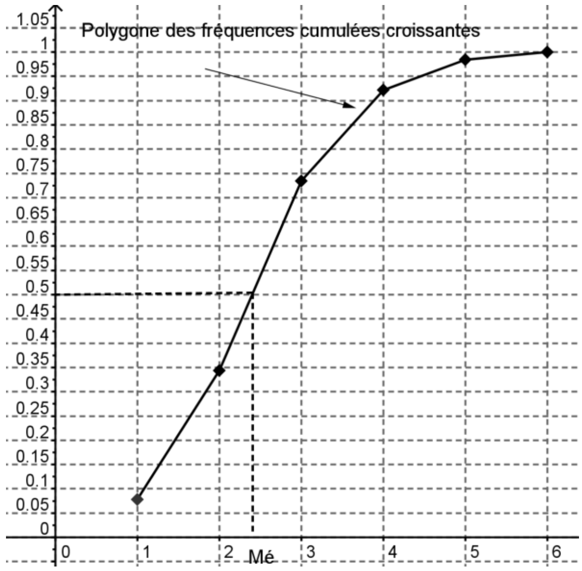
3) Les effectifs cumulés croissants de cette série

Nombre de pièces (x_i)	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Nombre d'appartement (n_i)	**10**	**34**	**50**	**24**	**8**	**2**	**N=128**
Effectifs cumulés croissants	**10**	**44**	**94**	**118**	**126**	**128=N**	

4) **a)** Les fréquences cumulées croissantes de cette série

Nombre des pièces (x_i)	**1**	**2**	**3**	**4**	**5**	**6**
Nombre d'appartements (n_i)	**10**	**34**	**50**	**24**	**8**	**2**
Fréquences F_i	0,078125	0,265625	0,390625	0,1875	0,0625	0,015625
Fréquences cumulées croissantes	0,078125	0,34375	0,734375	0,921875	0,984375	**F = 1**

b) Le polygone des fréquences cumulées croissantes correspondant



la médiane est l'abscisse du point de la courbe dont l'ordonnée est égale à **0,5** donc M_e □ 3

5) Moyenne : $\overline{x} = 2,9375$

6) Écart type : $\sigma = 1,073472752$

Variance V : $V = 1,15234375$

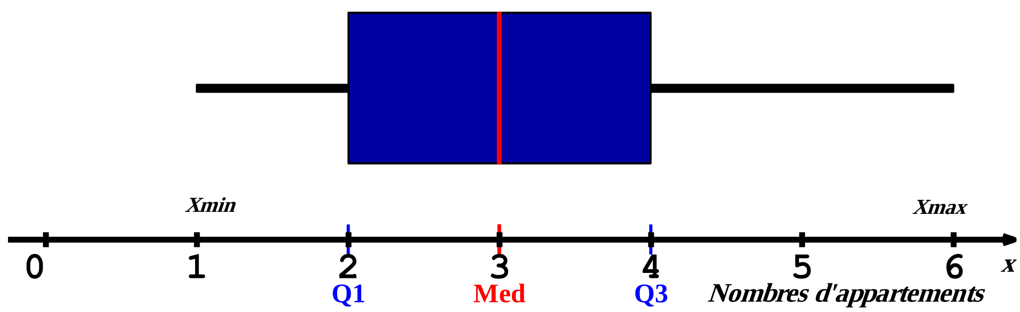
7) Les quartiles de cette série :

Q_1 **:** La valeur du caractère de l'individu de rang $\frac{N}{4} = \frac{128}{4} = 32$ (n_i □) donc $Q_1 = 2$

$Q_2 = M_é = 3$

Q_3 =La valeur du caractère de l'individu de rang $\frac{3N}{4} = 96$ (n_i □) donc $Q_3 = 4$

8) Le diagrammes en boite ou la boite de dispersion de cette série :

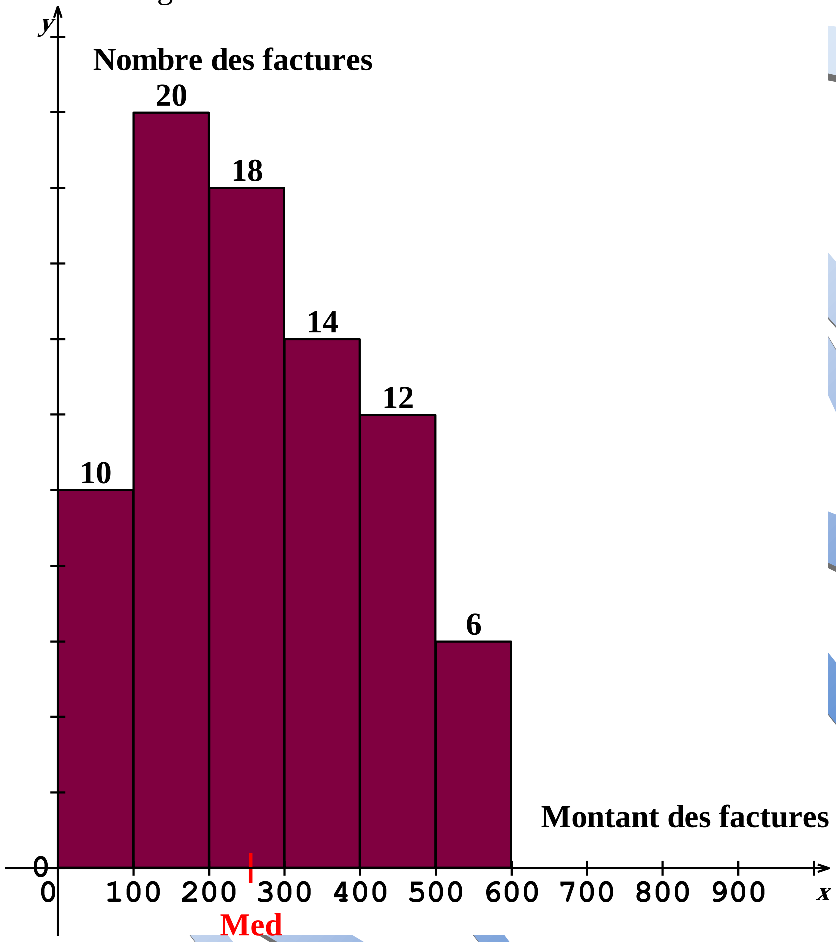


Activité N°3 : (Cas d'une série à caractère continue)

Un grossiste a relevé le montant des factures de livraison reçue au cours d'un mois.

Montant des factures (En dinars)	[0,100[	[100,200[	[200,300[	[300,400[	[400,500[	[500,600[
Nombre des factures	**10**	**20**	**18**	**14**	**12**	**6**

1. Le histogramme de cette série



2. a) La classe modale de cette série est [100,200[donc le mode de cette série est

$$M_o = \frac{100+200}{2} = 150$$

b) Le centre de chaque classe :

Montant des factures (En dinars)	[0,100[	[100,200[	[200,300[	[300,400[	[400,500[	[500,600[
Nombre des factures	**10**	**20**	**18**	**14**	**12**	**6**
centre de l'intervalle c_i	**50**	**150**	**250**	**350**	**450**	**550**

c) La moyenne $\overline{x}$ de cette série :

Moyenne : $\overline{x} = 270,063$; **calculatrice :** c_i STO n_i M^+

3. l'écart type σ et la variance V de cette série statistique :

Écart type : $\sigma = 146,876$

Variance V : $V = 2157,55938$

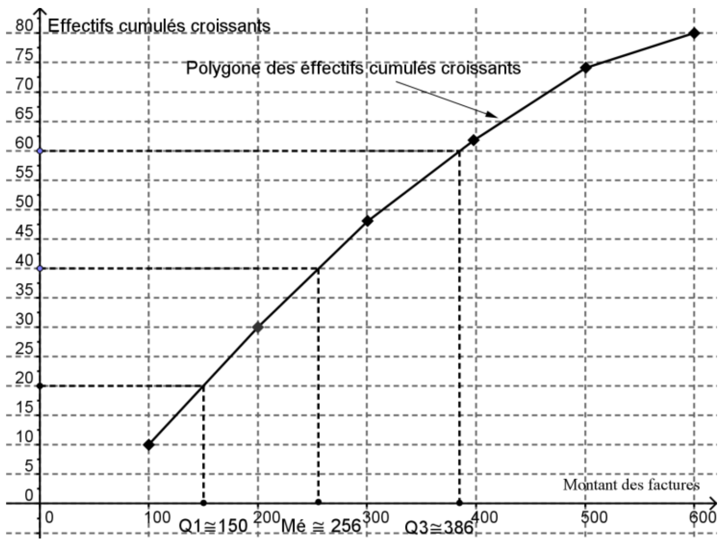
4. Les fréquences cumulées croissantes de cette série

Montant des factures (En dinars)	[0,100[	[100,200[	[200,300[	[300,400[	[400,500[	[500,600[	Total
Nombre des factures	**10**	**20**	**18**	**14**	**12**	**6**	**N=80**
Fréquences F_i	0,125	0,250	0,225	0,175	0,150	0,075	
Fréquences cumulées croissantes	0,125	0,375	0,6	0,775	0,925	**F = 1**	

5. a) Les effectifs cumulés croissants de cette série

Montant des factures (En dinars)	[0,100[	[100,200[	[200,300[	[300,400[	[400,500[	[500,600[	Total
Nombre des factures	**10**	**20**	**18**	**14**	**12**	**6**	**N=80**
Effectifs cumulés croissants	**10**	**30**	**48**	**62**	**74**	**N= 80**	

b) Le polygone des effectifs cumulés croissants de cette série



Le polygone des effectifs cumulés croissants correspondant
La moitié de l'effectif total est $\frac{N}{2}=\frac{80}{2}=40$ d'après le graphique, la médiane est l'abscisse du point de la courbe dont l'ordonnée est égale à40 donc le médiane $M_e \approx 255$

6. Les quartiles de cette série :

Graphiquement : sur le polygone des effectifs cumulés croissants

- Le point d'ordonnée est d'abscisse $\frac{N}{4}=\frac{80}{4}=20$ est d'abscisse Q_1 , on lit $Q_1 \approx 150$
- Le point d'ordonnée est d'abscisse $\frac{N}{2}=\frac{80}{2}=40$ est d'abscisse Q_2 , on lit $Q_2 = M_é \approx 256$
- Le point d'ordonnée est d'abscisse $\frac{3N}{4}=60$ est d'abscisse Q_3 ,on lit $Q_3 \approx 386$

Calcul des quartiles par interpolation linéaire : N = 80

- Q_1 appartient à la classe du caractère de l'individu de rang $\frac{N}{4}=20$ $(n_i \nearrow)$ donc $Q_1 \in [100,200[$

x_i	$n_i \nearrow$
100	**10**
Q_1 =?	$\frac{N}{4}=20$
200	**30**

$\frac{Q_1-100}{20-10}=\frac{200-100}{30-10}=5$ **d'ou** Q_1 **=150**

❖ Q_2 appartient à la classe du caractère de l'individu de rang $\frac{N}{2} = 40$ (n_i) donc $Q_2 \in [200, 300[$

x_i	n_i
200	**30**
Q_2 =?	$\frac{N}{2} = 40$
300	**48**

$$\frac{Q_2 - 200}{40 - 30} = \frac{300 - 200}{48 - 30} = \frac{100}{18} \textbf{ donc } Q_2 = \frac{1000}{18} + 200 \approx 256$$

❖ Q_3 appartient à la classe du caractère de l'individu de rang $\frac{3N}{4} = 60$ (n_i) donc $Q_3 \in [300, 400[$

x_i	n_i
300	**48**
Q_3 =?	$\frac{3N}{4} = 60$
400	**62**

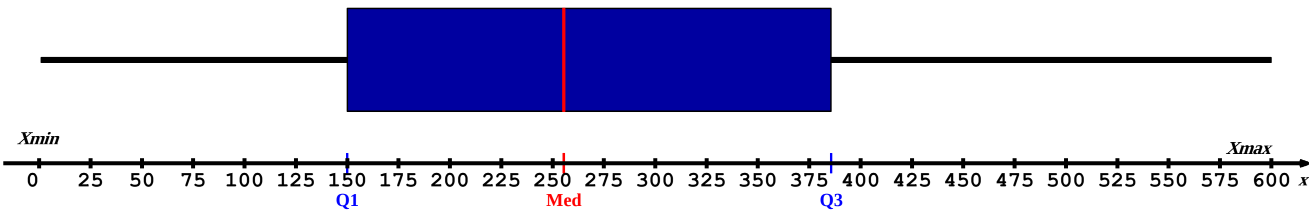
$$\frac{Q_3 - 300}{60 - 48} = \frac{400 - 300}{62 - 48} = \frac{100}{14} \textbf{ donc } Q_3 = \frac{1200}{14} + 300 \approx 386$$

7. Le diagrammes en boite ou la boite de dispersion de cette série :

Minimum : 0 ; Maximum : 600

Premier quartile : Q_1 = 150 ; Médiane : Q_2 = 256

Troisième quartile : Q_3 =386 ; Intervalle interquartile : [150, 386]



II) Série statistique double:

Au cours d'enquêtes ou d'expériences de laboratoire, on est parfois amené à étudier une population selon deux caractères à la fois.

Exemple 1:

Le tableau ci-dessous représente le poids en kg et la taille en m des joueurs d'une équipe de basket-ball.

Poids en kg (x_i)	75	80	85	90	95	100	102	105	110	115
Taille en m (y_i)	1 ,75	1,85	1,90	1,95	2,00	2 ,05	2 ,07	2 ,08	2 ,09	2 ,10

Le couple (x_1 , y_1) = (75 , 1,75) veut dire que le joueur N°1 pèse 75 Kg et mesure 1 ,75 m.

On a donc une population de 10 joueurs sur laquelle on a observé simultanément les deux variables x_i et y_i.

1) Définition :

On dit qu'un couple (X, Y) de variables statistiques définies une série double si les deux Variables X et Y sont observés simultanément sur une même population.

2) Nuage de points associé à une série statistique double – Point moyen d'un nuage

Activité N°1 :

Le tableau ci-dessous représente le poids en kg et la taille en m des joueurs d'une équipe de basket-ball.

Poids (x_i)	75	80	85	90	95	100	102	105	110	115
Taille (y_i)	1 ,75	1,85	1,90	1,95	2,00	2 ,05	2 ,07	2 ,08	2 ,09	2 ,10

1. *Placer dans un repère orthogonal $(o,\vec{i},\vec{j})$ l'ensemble des points Mi (x_i , y_i)*
2. *Appelons $\overline{X}$ (respectivement $\overline{Y}$) la moyenne arithmétique de la série statistique à variable le poids en kg (respectivement la taille en m)*
   - b) *Calculer $\overline{X}$ et $\overline{Y}$*
   - c) *Placer le point G($\overline{X}$, $\overline{Y}$) dans le repère $(o,\vec{i},\vec{j})$.*

Définition :

Soit une série statistique définie par deux variables X et Y . On désigne par x_1, x_2 ,......, x_n les valeurs de X et par y_1, y_2 ,......, y_n celles de Y. Le plan étant rapporté à un repère orthogonal.

- *L'ensemble des points Mi (x_i , y_i) ; $i \in$ {1,2,.....,n} est appelé le nuage de points associé à la série statistique double.*
- *Le point G($\overline{X}$, $\overline{Y}$) est appelé point moyen de ce nuage.* $\overline{X} = \sum_{i=1}^{n} \frac{x_i}{n}$ **et** $\overline{Y} = \sum_{i=1}^{n} \frac{y_i}{n}$

  $\overline{X}$ *la moyenne arithmétique des réels* x_i

  $\overline{Y}$ *la moyenne arithmétique des réels* y_i

Utilisation des calculatrices

Série statistique individuelle :

Le tableau ci-dessous représente le poids en kg et la taille en m des joueurs d'une équipe de basket-ball.

Poids (x_i)	**75**	**80**	**85**	**90**	**95**	**100**	**102**	**105**	**110**	**115**
Taille (y_i)	**1,75**	**1,85**	**1,90**	**1,95**	**2,00**	**2,05**	**2,07**	**2,08**	**2,09**	**2,10**

Pour choisir le mode de fonctionnement en statistique appuyer sur 2ndF MODE 2

Pour entrer les données taper

x_i	STO	y_i	M^+
75	STO	**1,75**	M^+
80	STO	**1,85**	M^+
85	STO	**1,90**	M^+
90	STO	**1,95**	M^+
95	STO	**2,00**	M^+
100	STO	**2,05**	M^+
102	STO	**2,07**	M^+
105	STO	**2,08**	M^+
110	STO	**2,09**	M^+
115	STO	**2,10**	M^+

La calculatrice affiche l'effectif total n=10

Pour afficher la moyenne $\overline{X}$ appuyer sur :	RCL	$\overline{X}$	*On obtient :*	**$\overline{X}$ =95,7**
Pour afficher la moyenne $\overline{Y}$ appuyer sur :	RCL	$\overline{Y}$	*On obtient :*	**$\overline{Y}$ =1,983**
Pour afficher L'écart type σ_x appuyer sur :	RCL	σ_x	*On obtient :*	**σ_x =12, 42618204**
Pour afficher L'écart type σ_y appuyer sur :	RCL	σ_y	*On obtient :*	**σ_y =0, 111628849**
Pour afficher la variance V(x) appuyer sur :	RCL	σ_x x^2	*On obtient :*	**V(x) = 154,41**
Pour afficher la variance V(y) appuyer sur :	RCL	σ_y x^2	*On obtient :*	**V(y) = 0,012461**

Distribution marginale.

1) Définition : *Soient deux variables statistique définies sur une population d'effectif total N. On désigne par x_1, x_2 ,......, x_p les valeurs de X et par y_1, y_2 ,......, y_q celles de Y.*

Le nombre d'individus de la population vérifiant simultanément X= x_i et Y= y_i est noté n_{ij} associés aux couples (x_i , y_i) sont présentés à l'aide d'un tableau à double entrée de la forme :

Y / **X**	y_1	y_2	....	y_j	......	y_q	**Distribution marginale de X**
x_1	n_{11}	n_{12}	......	n_{1j}	......	n_{1q}	$n_{1.}$
x_2	n_{21}	n_{22}	......	n_{2j}	......	n_{2q}	$n_{2.}$
....	......	......	......	......	......	......	....
x_i	n_{i1}	n_{i2}	......	n_{ij}	......	n_{iq}	$n_{i.}$
......	......	......	......	......	......	......	......
x_p	n_{p1}	n_{p2}	......	n_{pj}	......	n_{pq}	$n_{p.}$
Distribution marginale de Y	$n_{.1}$	$n_{.2}$	......	$n_{.1}$	....	$n_{.q}$	$N=\sum_{i=1}^{p} n_{i.}=\sum_{j=1}^{q} n_{.j}$

2) Fréquences marginales :

Définition:

Les rapports des effectifs d'une distribution marginale d'un caractère à l'effectif totale définissent les fréquences marginales de ce caractère.

- *Les fréquences marginales du caractère X*

La distribution marginale de la variable X est la distribution des différentes valeurs prises par la variable X.

X	x_1	x_2	…….	x_i	……	x_p	**Total**
Effectif	$n_{1.}$	$n_{2.}$	……	$n_{i.}$	……	$n_{p.}$	$N=\sum_{i=1}^{p} n_{i.}$
Fréquence	$F_1=\frac{n_{1.}}{N}$	$F_2=\frac{n_{2.}}{N}$	……	$F_i=\frac{n_{i.}}{N}$	……	$F_P=\frac{n_{P.}}{N}$	

- *Le fréquences marginales du caractère Y*

La distribution marginale de la variable Y est la distribution des différentes valeurs prises par la variable Y.

Y	x_1	y_2	…….	y_j	……	y_q	**Total**
Effectif	$n_{.1}$	$n_{.2}$	……	$n_{.j}$	……	$n_{.q}$	$N=\sum_{j=1}^{q} n_{.j}$
Fréquence	$F_1=\frac{n_{.1}}{N}$	$F_2=\frac{n_{.2}}{N}$	……	$F_i=\frac{n_{.j}}{N}$	……	$F_P=\frac{n_{.q}}{N}$	

3) Moyenne - Variance V - Ecart type :

$$\overline{X}=\frac{\sum_{i=1}^{p} n_i x_i}{N}\ ,\quad V(X)=\frac{n_1(x_1)^2+n_2(x_2)^2+.....+n_p(x_p)^2}{N}-(\overline{X})^2=\frac{\sum_{i=1}^{p} n_i x^2{}_i}{N}-(\overline{X})^2,\quad \sigma(X)=\sqrt{V(X)}$$

$$\overline{Y}=\frac{\sum_{i=1}^{q} n_i y_i}{N}\ ,\quad V(Y)=\frac{n_1(y_1)^2+n_2(y_2)^2+.....+n_q(y_q)^2}{N}-(\overline{Y})^2=\frac{\sum_{i=1}^{q} n_i x^2{}_i}{N}-(\overline{Y})^2,\quad \sigma(Y)=\sqrt{V(Y)}$$

Activité N°2 :

On a étudie la répartition des 20 appartements d`un immeubles suivant le nombre $x_i$ de personnes et le nombre $y_i$ de pièces d`habitation.

X \ Y	2	3	4
2	1	3	0
3	0	2	1
4	1	4	3
5	2	1	2

1) *Déterminer les distribution des effectifs marginaux et de fréquences marginales de chacune des variables X et de Y .*
2) *a)Calculer $\overline{X}$ et $\overline{Y}$ moyenne respectives des variables statistiques X et Y .*
   b) Calculer les écart-types $\sigma(X)$ et $\sigma(Y)$ et les variances $V(X)$ et $V(Y)$ de cette série statistique
3) *a) Construire le nuage de points de cette série*
   b) Placer le point moyen G du nuage

Utilisation des calculatrices

Série statistique à double entrée :

On a étudie la répartition des 20 appartements d`un immeubles suivant le nombre $x_i$ de personnes et le nombre $y_i$ de pièces d`habitation.

Y / X	2	3	4
2	1	3	0
3	0	2	1
4	1	4	3
5	2	1	2

Même procédé que les séries individuelles

Pour choisir le mode de fonctionnement en statistique appuyer sur $\boxed{\text{2ndF}}$ $\boxed{\text{MODE}}$ $\boxed{2}$

Pour entrer les données taper

x_i	STO	y_i	STO	n_{ij}		M^+
2	STO	**2**	STO	**1**		M^+
3	STO	**2**	STO	**0**		M^+
4	STO	**2**	STO	**1**		M^+
5	STO	**2**	STO	**2**		M^+
2	STO	**3**	STO	**3**		M^+
3	STO	**3**	STO	**2**		M^+
4	STO	**3**	•••	••••		•••
5	STO	**3**	•••	••••		••••
2	STO	**4**	•••	••••		••••
3	STO	**4**	••••	••••		••••
•••	••••	•••	••••	••••		••••
•••	••••	•••	••••	••••		••••

Pour afficher la moyenne $\overline{X}$ appuyer sur :	RCL	$\overline{X}$		On obtient :	$\overline{X}$ =
Pour afficher la moyenne $\overline{Y}$ appuyer sur :	RCL	$\overline{Y}$		On obtient :	$\overline{Y}$ =
Pour afficher L'écart type σ_x appuyer sur :	RCL	σ_x		On obtient :	σ_x =
Pour afficher L'écart type σ_y appuyer sur :	RCL	σ_y		On obtient :	σ_y =
Pour afficher la variance V(x) appuyer sur :	RCL	σ_x	x^2	On obtient :	**V(x)** =
Pour afficher la variance V(y) appuyer sur :	RCL	σ_y	x^2	On obtient :	**V(y)** =

Ajustement affine.

Soit (X , Y) une série statistique double et G son point moyen.
On divise le nuage de point de (X , Y) en deux partie contenant à peu prés le même nombre de points obtenant ainsi deux nuages de points. On désigne par G1 et G2 les points moyens de ces deux nuages.
La droite (G1G2) passe par le point G est définit un ajustement affine du nuage de points représentant la série statistique double (X , Y).

Interprétation graphique.

Si le nuage de points représentant la série statistique double (X , Y) a une forme allongée et si les points qui le forment se repartissent sensiblement autour d'une droite passant par G, cela permet de dire que l'on peut lier Y et X à l'aide d'une fonction affine.

***Exemple N°1** :*

Des élèves effectuent diverses mesures de la longueur d'un ressort pour différente masses accrochées à ce ressort.

X (g) 100 200 300 500 1000 1500 2000

Y (cm) 20,3 20,8 21,2 22 23,9 26 28

1°) Représenter le nuage de points M(X , Y) dans un repère orthogonal.

2°) Placer sur la figure le point G(x , y)

3°) En examinant le nuage de points, peut-on conclure qu'il existe une relation entre Y et X ? de quelle type ?

9 Statistiques

9-1 Séries statistiques simples

• Moyenne, Variance et écart type d'une série statistique :
(valeurs du caractère : x_i ; effectif : n_i ; effectif total : N)
Moyenne : $\bar{x} = \dfrac{n_1x_1 + n_2x_2 + n_3x_3 + \cdots}{N}$
Variance $V(x) = \dfrac{n_1(x_1)^2 + n_2(x_2)^2 + n_3(x_3)^2 + \cdots}{N} - (\bar{x})^2$; Ecart type : $\sigma(x) = \sqrt{V(x)}$

► *Exemple :*

x_i	1	2	3	4	5
n_i	10	8	1	3	1

• $\bar{x} = \dfrac{10 \times 1 + 8 \times 2 + 1 \times 3 + 3 \times 4 + 1 \times 5}{10 + 8 + 1 + 3 + 5} = 2$

• $V(x) = \dfrac{10 \times 1^2 + 8 \times 2^2 + 1 \times 3^2 + 3 \times 4^2 + 1 \times 5^2}{10 + 8 + 1 + 3 + 5} - 2^2 = \dfrac{32}{23}$; $\sigma(x) = \sqrt{V(x)} \approx 1,18$

9-2 Séries statistiques doubles

Pour une série :

Caractère x_i	x_1	x_2	$\cdots$	x_n
Caractère y_i	y_1	y_2	$\cdots$	y_n

• Point moyen : $G\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$ ($\bar{x}$: moyenne des x_i ; $\bar{y}$: moyenne des y_i)

• Covariance de x et y : $C_{xy} = \dfrac{1}{n}(x_1y_1 + x_2y_2 + \cdots + x_ny_n) - \bar{x}\bar{y}$

• Droite des moindres carrés : $y = ax + b$ avec $a = \dfrac{C_{xy}}{V(x)}$ et $b = \bar{y} - a\bar{x}$

► *Exemple :*

x_i	1	2	3	4	5
y_i	8	9	12	12	14

• $\bar{x} = \dfrac{1 + 2 + 3 + 4 + 5}{5} = 3$; $\bar{y} = \dfrac{8 + 9 + 12 + 12 + 14}{5} = 11$; $G\begin{pmatrix} 3 \\ 11 \end{pmatrix}$

• $C_{xy} = \dfrac{1}{5}(1 \times 8 + 2 \times 9 + 3 \times 12 + 4 \times 12 + 5 \times 14) - 3 \times 11 = 3$

• $V(x) = \dfrac{1^2 + 2^2 + 3^2 + 4^2 + 5^2}{5} - 3^2 = 2$

• Droite des moindres carrés : $a = \dfrac{C_{xy}}{V(x)} = \dfrac{3}{2} = 1,5$ et $b = \bar{y} - a\bar{x} = 11 - 1,5 \times 3 = 6,5$
Une équation de la droite des moindres carrés est donc : $y = 1,5x + 6,5$

• Estimation de la valeur de y pour $x = 7$: $y = 1,5 \times 7 + 6,5 = 17$

N*MY